table extraction

Mingze Li

There are many tools to extract table from pdf files, we will compare them and choose the one that fits our use case best.

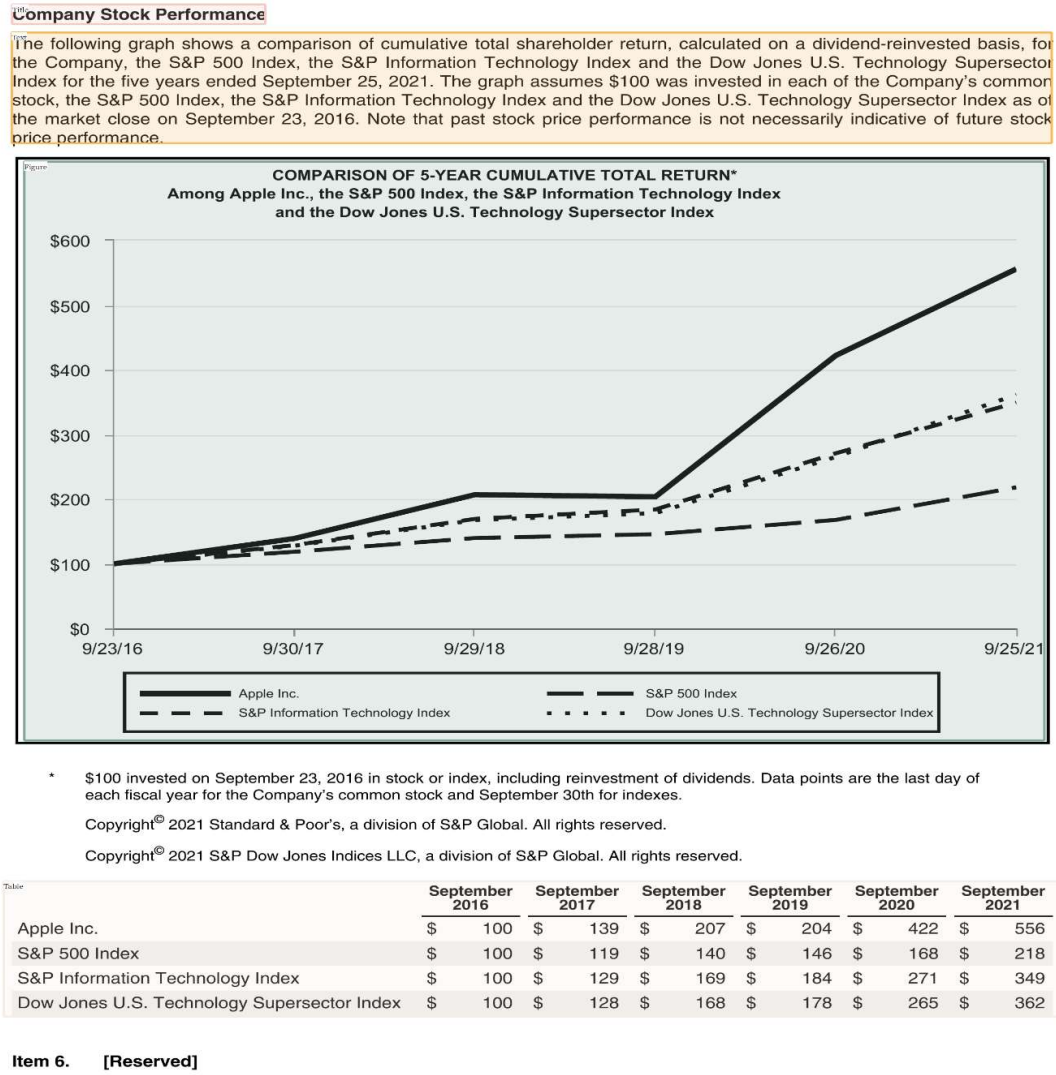
(Here is all the setup, I use Colab for all the test except wxTableExtract)

```
[ ] !pip install pdf2image
    !apt-get install poppler-utils
    !apt-get install libpoppler-cpp-dev
    !pip install python-poppler
    !pip3 install torch torchvision torchaudio --extra-index-url https://download.pytorch.org/whl/cu116
    !pip install layoutparser torchvision
    !pip install "git+https://github.com/facebookresearch/detectron2.git@v0.5#egg=detectron2"
    !pip install "layoutparser[ocr]"
    !pip install tabula-py
    !sudo apt-get update
```

```
import warnings
warnings.filterwarnings("ignore")

## for data
import os
import io
import pandas as pd
import numpy as np

## for plotting
import matplotlib.pyplot as plt

## for ocr
import pdf2image
import cv2
import layoutparser as lp
from PIL import Image
import os

import tabula

from google.colab import drive
drive.mount('/content/drive')
os.chdir("/content/drive/MyDrive/project")
```

We start with OCR. Here we use layoutparser as the CV engine. It contains several deep learning based pre-trained models for detection. For our tests, we chose PubLayNet to detect text, titles, lists, tables, and figures in the picture and then enclosing it with a rectangular bounding box.

```
doc = pdf2image.convert_from_path("doc_apple.pdf", dpi=300)
## load pre-trained model
model = lp.Detectron2LayoutModel(
      "lp://PubLayNet/mask_rcnn_X_101_32x8d_FPN_3x/config",
      extra_config=["MODEL.ROI_HEADS.SCORE_THRESH_TEST", 0.8],
      label_map={0:"Text", 1:"Title", 2:"List", 3:"Table", 4:"Figure"})
## turn img into array
i = 21
img = np.asarray(doc[i])
## predict
detected = model.detect(img)
## plot
lp.draw_box(img, detected, box_width=5, box_alpha=0.2,
                    show_element_type=True)
```



Company Stock Performance

The following graph shows a comparison of cumulative total shareholder return, calculated on a dividend-reinvested basis, for the Company, the S&P 500 Index, the S&P Information Technology Index and the Dow Jones U.S. Technology Supersector Index for the five years ended September 25, 2021. The graph assumes $100 was invested in each of the Company's common stock, the S&P 500 Index, the S&P Information Technology Index and the Dow Jones U.S. Technology Supersector Index as of the market close on September 23, 2016. Note that past stock price performance is not necessarily indicative of future stock price performance.

* $100 invested on September 23, 2016 in stock or index, including reinvestment of dividends. Data points are the last day of each fiscal year for the Company's common stock and September 30th for indexes.

Copyright© 2021 Standard & Poor's, a division of S&P Global. All rights reserved.

Copyright© 2021 S&P Dow Jones Indices LLC, a division of S&P Global. All rights reserved.

	September 2016	September 2017	September 2018	September 2019	September 2020	September 2021
Apple Inc.	$ 100	$ 139	$ 207	$ 204	$ 422	$ 556
S&P 500 Index	$ 100	$ 119	$ 140	$ 146	$ 168	$ 218
S&P Information Technology Index	$ 100	$ 129	$ 169	$ 184	$ 271	$ 349
Dow Jones U.S. Technology Supersector Index	$ 100	$ 128	$ 168	$ 178	$ 265	$ 362

Item 6. [Reserved]

The output shows that it can detect the information correctly.

Because our focus is on the extraction of tables, we try to extract the information in this table by layoutparser[ocr].

```
## sort

new_detected = detected.sort(key=lambda x: x.coordinates[1])
## assign ids
detected = lp.Layout([block.set(id=idx) for idx,block in
                                        enumerate(new_detected)])
'''
{'0-Title': '...',
 '1-Text':  '...',
 '2-Figure': array([[ [0,0,0], ...]]),
 '3-Table': pd.DataFrame,
}
'''
def parse_doc(dic):
        for k,v in dic.items():
                if "Title" in k:
                        print('\x1b[1;31m'+ v +'\x1b[0m')
                elif "Figure" in k:
                        plt.figure(figsize=(10,5))
                        plt.imshow(v)
                        plt.show()
                else:
                        print(v)
                print(" ")
model = lp.TesseractAgent(languages='eng')
dic_predicted = {}
for block in [block for block in detected if block.type == "Table"]:
        ## segmentation
        segmented = block.pad(left=15, right=15, top=5,
                              bottom=5).crop_image(img)
        ## extraction
        extracted = model.detect(segmented)
        ## save
        dic_predicted[str(block.id)+"-"+block.type] = pd.read_csv(
                                io.StringIO(extracted) )
# check
parse_doc(dic_predicted)
```

	September September September September September September
0	2016 2017 2018 2019 2020 2021
1	Apple Inc. 100 139 207 204 422 556
2	S&P 500 Index 100 119 140 146 168 218
3	S&P Information Technology Index 100 129 $...
4	Dow Jones U.S. Technology Supersector Index $ ...

We can see the extraction isn't great, it can't recognize the columns correctly.
So, we try the second tools: tabula.

```
tables = tabula.read_pdf("doc_apple.pdf", pages=i+1)
tables[0]
```

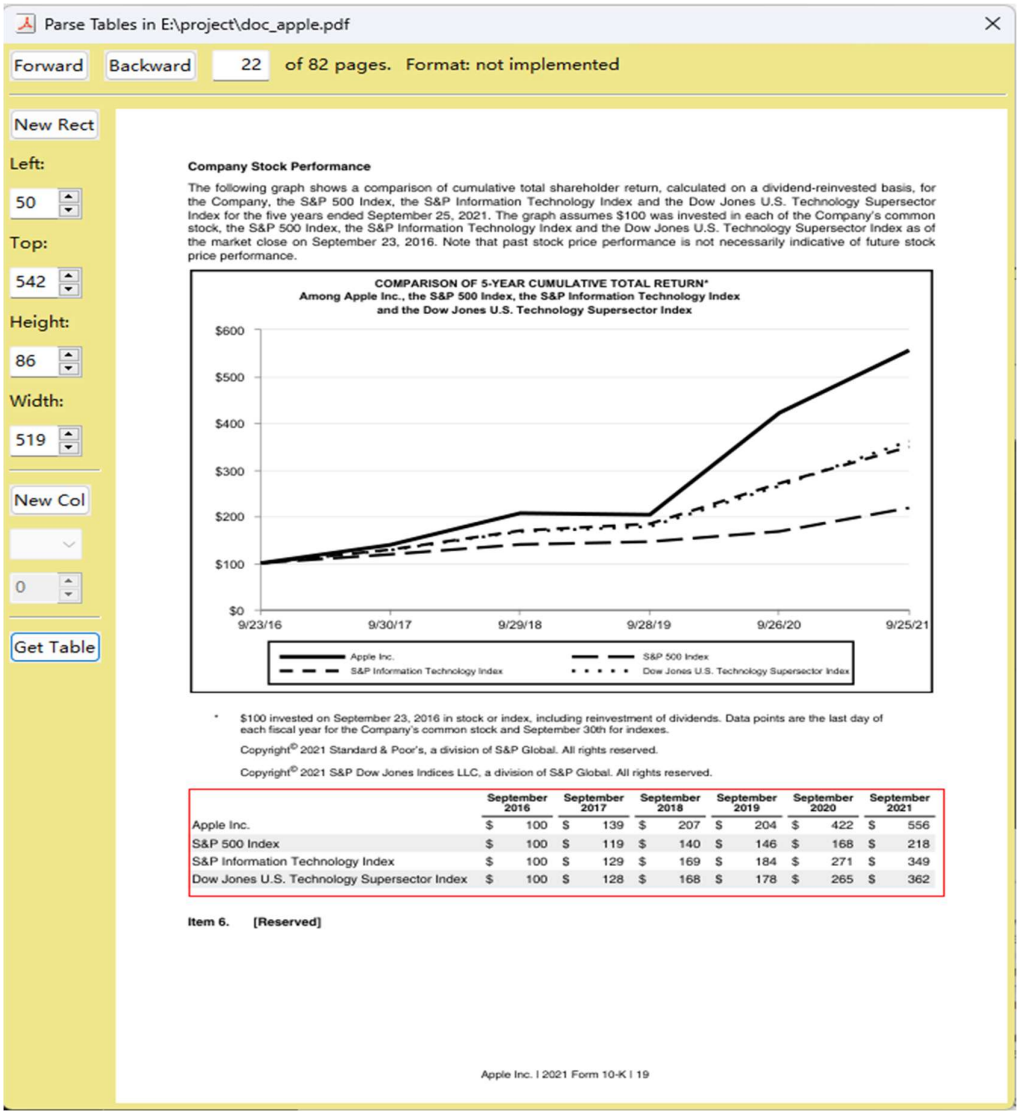
	Unnamed: 0	September	September.1	September.2	September.3	September.4	September.5
0	NaN	2016	2017	2018	2019	2020	2021
1	Apple Inc.	$ 100	$ 139	$ 207	$ 204	$ 422	$ 556
2	S&P 500 Index	$ 100	$ 119	$ 140	$ 146	$ 168	$ 218
3	S&P Information Technology Index	$ 100	$ 129	$ 169	$ 184	$ 271	$ 349
4	Dow Jones U.S. Technology Supersector Index	$ 100	$ 128	$ 168	$ 178	$ 265	$ 362

As you can see, even if the names are still wrong, it can recognize the columns now.

How about wxTableExtract?

I've downloaded the repository from: https://github.com/pymupdf/PyMuPDF-Utilities

Since this program needs desktop to show a GUI on, I use PyCharm to run it.

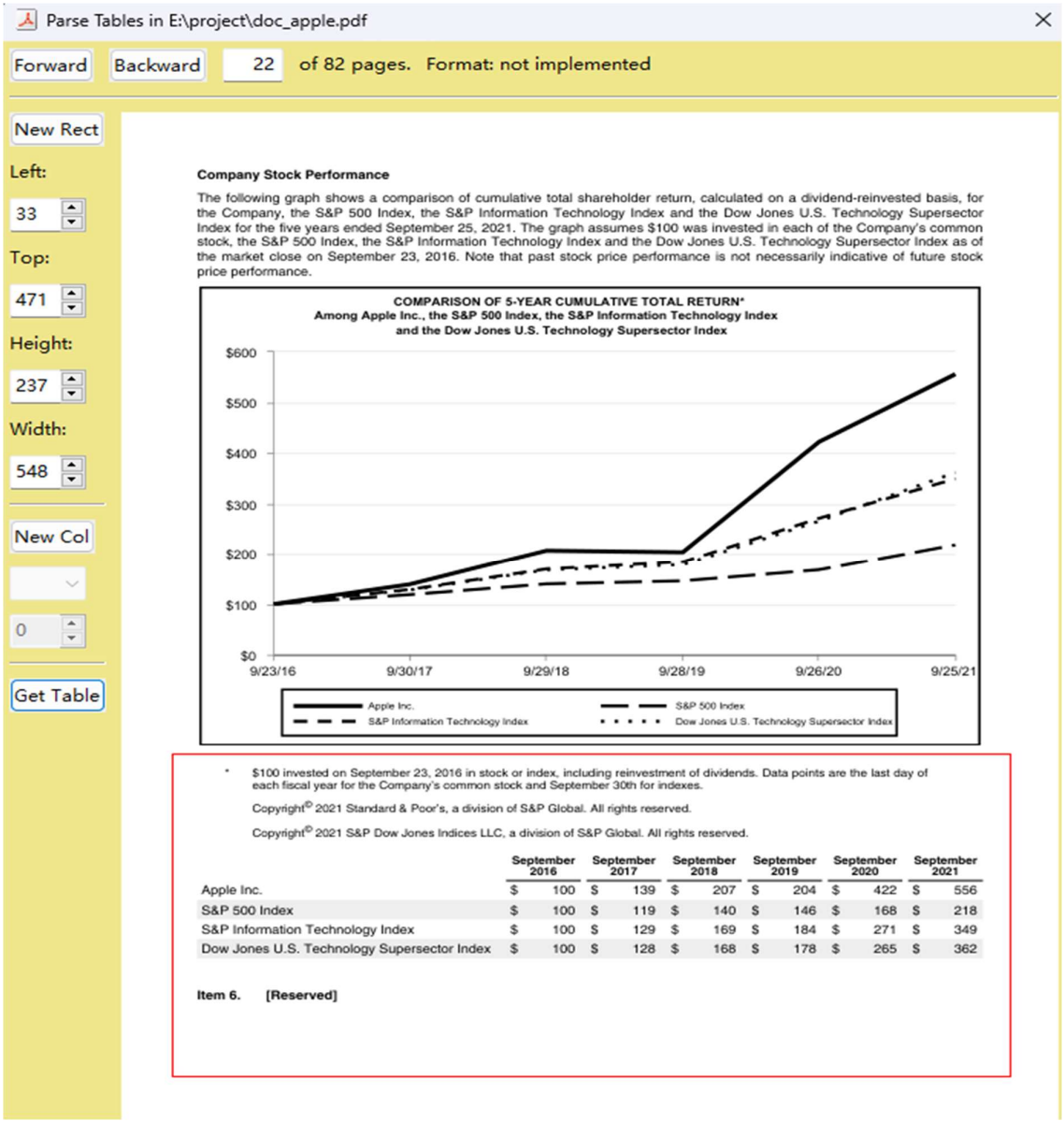


Here is the output:

```
C:\Users\15392\anaconda3\envs\machinelearning\python.exe E:\project\PyMuPDF-Utilities-master\examples\wxTableExtract.py

Contents of (6 x 1)-table at [50,542] on page 22
['September September September September September September']
['2016 2017 2018 2019 2020 2021']
['Apple Inc. $ 100 $ 139 $ 207 $ 204 $ 422 $ 556']
['S&P 500 Index $ 100 $ 119 $ 140 $ 146 $ 168 $ 218']
['S&P Information Technology Index $ 100 $ 129 $ 169 $ 184 $ 271 $ 349']
['Dow Jones U.S. Technology Supersector Index $ 100 $ 128 $ 168 $ 178 $ 265 $ 362']

Process finished with exit code 0
```

It can't recognize the columns and can't read the name correctly neither.

Meanwhile, it cannot detect the table automatically. If we draw a larger rectangle which include other information, wxTableExtract won't be able to detect the table in the rectangle.




```
C:\Users\15392\anaconda3\envs\machinelearning\python.exe E:\project\PyMuPDF-Utilities-master\examples\wxTableExtract.py

Contents of (13 x 1)-table at [33,471] on page 22
['* $100 invested on September\xa023, 2016 in stock or index, including reinvestment of dividends. Data points are the last day of']
['each fiscal year for the Company's common stock and September 30th for indexes.']
['Copyright©']
['2021 Standard & Poor's, a division of S&P Global. All rights reserved.']
['Copyright©']
['2021 S&P Dow Jones Indices LLC, a division of S&P Global. All rights reserved.']
['September September September September September September']
['2016 2017 2018 2019 2020 2021']
['Apple Inc. $ 100 $ 139 $ 207 $ 204 $ 422 $ 556']
['S&P 500 Index $ 100 $ 119 $ 140 $ 146 $ 168 $ 218']
['S&P Information Technology Index $ 100 $ 129 $ 169 $ 184 $ 271 $ 349']
['Dow Jones U.S. Technology Supersector Index $ 100 $ 128 $ 168 $ 178 $ 265 $ 362']
['Item 6. [Reserved]']
```

It seems that Tabula perform best when extracting table.

But “doc_apple.pdf” is just a simple test which is clearly written and very readable. To see how tabula would behave on a more complex report, we tried "BMO-2019-Sustainability-Report-PAS-2.pdf" as input.

```
[ ] tables_raw = tabula.read_pdf("BMO-2019-Sustainability-Report-PAS-2.pdf", pages="all")
    for table in tables_raw:
        display(table)    # <-- visualize a page
```

	Unnamed: 0	development programs; and our community	relationships with our customers, including
0	BMO’ s Purpose Manifesto	giving and engagement with our partners.	women entrepreneurs and small
1	NaN	NaN	businesses, to address the challenge of
2	Since the beginning, BMO has had a purpose bey...	To deliver on and amplify our Purpose,	economic disparity.
3	for progress and a catalyst for change. We are...	we will:	NaN
4	seeking to fulfill their dreams, we are an emp...	NaN	• Our employees: championing diversity
5	NaN	• Work to increase positive impacts and	and inclusion across our organization and
6	of people who work with us every day, and we a...	reduce negative impacts across BMO’ s	preparing our colleagues for the jobs of
7	communities in which we live and work. Our com...	operations and value chain.	the future.
8	change for our customers, employees and commun...	NaN	NaN
9	NaN	• Develop new products and services to	• Our environment: managing our
10	than 200 years. In short, our Purpose is to Bo...	NaN	NaN
11	NaN	help our customers achieve society’ s	environmental footprint and financing
12	and life.	shared economic, social and	and advising our clients during the
13	NaN	environmental objectives.	transition to a resource-efficient, low-
14	NaN	NaN	carbon and circular economy.
15	NaN	• Embed sustainability more deeply within	NaN
16	Our success has involved adapting to a In supp...	our operations and internal procedures.	• Our community: building stronger,
17	changing society while helping our customers c...	NaN	healthier communities through donations,
18	do the same. Today, from human rights to	• Support collaboration, learning and	sponsorships and volunteering.
19	climate change to sustainable finance, the • F...	innovation through our community	NaN
20	challenges we face are increasingly complex mo...	giving activities.	• Our suppliers: dealing fairly and ethically
21	and interconnected. In response, we are financ...	NaN	with our suppliers and working with them
	strengthening our commitment to fund to find	• Actively engage stakeholders and	

We asked tabula to extract the tables from the first 20 pages and observed that the first table was very strange.

BMO's Purpose Manifesto

Since the beginning, BMO has had a purpose beyond profit to be a champion for progress and a catalyst for change. We are a partner for our customers seeking to fulfill their dreams, we are an employer for the tens of thousands of people who work with us every day, and we are proud members of the communities in which we live and work. Our commitment to driving positive change for our customers, employees and communities has spanned more than 200 years. In short, our Purpose is to Boldly Grow the Good *in business and life*.

Our success has involved adapting to a changing society while helping our customers do the same. Today, from human rights to climate change to sustainable finance, the challenges we face are increasingly complex and interconnected. In response, we are strengthening our commitment to sustainability and to navigating these changes together with our stakeholders.

We believe that profit and progress go hand in hand. We will think big and pursue our growth objectives confidently, balancing risk and opportunity to create value that is shared with the world around us. Our Purpose encourages us to do the right thing every day.

In support of our Purpose, BMO has committed to **double the good** by 2025:

- **For a sustainable future:** we will mobilize $400 billion for sustainable finance, and launch an impact investment fund to find and scale solutions to sustainability challenges.
- **For a thriving economy:** we will double BMO's support for small businesses and for women entrepreneurs.
- **For an inclusive society:** we will reduce barriers to inclusion for employees, customers and stakeholders partnering with BMO.

Inspired by the United Nations Sustainable Development Goals (SDGs) and grounded in BMO's values of integrity, empathy, diversity and responsibility, these commitments will shape everything we do: the way we serve our customers and the products and services we design for them; our ongoing hiring practices and employee training and development programs; and our community giving and engagement with our partners.

To deliver on and amplify our Purpose, we will:

- Work to increase positive impacts and reduce negative impacts across BMO's operations and value chain.
- Develop new products and services to help our customers achieve society's shared economic, social and environmental objectives.
- Embed sustainability more deeply within our operations and internal procedures.
- Support collaboration, learning and innovation through our community giving activities.
- Actively engage stakeholders and investee companies to drive best practices.

BOLDLY GROW THE GOOD

IN BUSINESS AND LIFE

We will continue to focus our efforts in a where we can have a measurable impa

- **Our shareholders:** generating long-te value for our shareholders and engagi with them openly.
- **Our customers:** fostering inclusive relationships with our customers, inclu women entrepreneurs and small businesses, to address the challenge o economic disparity.
- **Our employees:** championing diversit and inclusion across our organization a preparing our colleagues for the jobs c the future.
- **Our environment:** managing our environmental footprint and financing and advising our clients during the transition to a resource-efficient, low-carbon and circular economy.
- **Our community:** building stronger, healthier communities through donati sponsorships and volunteering.
- **Our suppliers:** dealing fairly and ethic with our suppliers and working with th to improve their practices.
- **Our peers:** collaborating on thought leadership and innovation initiatives a leveraging our combined influence to create positive and sustainable change

Together, BMO's operating groups are mobilizing to make real progress towar these goals. By working together, we ca all elevate our performance and be the change that grows the good for us all.

After comparing with the original pdf file, we found that it is because Tabula identifies the dividing line as the beginning of a table and tries to extract it.

On page 13, we found a clear table

Sustainability Scorecard

Indicator definitions are listed in the Glossary of Terms on pages 108–111.

Priority Topics and Associated Metrics	2019[1]	2018[1]	2017[1]	Target (if applicable)
Access to Banking				
Number of women-owned businesses – Canada (total #)	109,377[2]	–	–	
Loans to women-owned businesses – Canada ($ millions)	8,510[3]	–	–	$3 billion by June 2021
Business Conduct				
Ethics training (% completion)▲	99.7	99.7	99.7	97%[4]
Significant fines or monetary sanctions▲	0	0	0	
Contributing to Communities				
Donations – Canada and United States ($ millions)▲	67.2	63.5	62.3	
Employee volunteer hours	18,394	19,500	17,000	
Total raised through BMO Employee Giving Campaign ($ millions)	22.8	22.3	21.6	
Corporate Governance				
Average tenure of members of Board of Directors (years)	8	8	8	
Women on Board of Directors (% of independent directors)	35.7	35.7	41.7	No less than ⅓
Customer Experience				
Average time to complete Banking reviews by BMO's Office of the Ombudsman (days)▲	37	37	63	
Average time to complete Investment reviews by BMO's Office of the Ombudsman (days)▲	41	43	95	
Customer loyalty surveys completed	1,168,955	350,937	301,094	
Banking reviews completed by BMO's Office of the Ombudsman (total #)▲	350	250	247	
Investment reviews completed by BMO's Office of the Ombudsman (total #)▲	52	41	56	
Net Promoter Score® (NPS) – enterprise-wide	52	47[5]	36	
Cyber Security, Information Security and Privacy				
Substantiated complaints regarding breaches of customer privacy and losses of customer data (total #)	4	9[6]	6	

Notes
1. Fiscal year (November 1 to October 31), unless otherwise indicated.
2. As at September 30.
3. As at July 31.
4. Based on normal employee turnover, a completion rate of 97% or above is considered to include 100% of the employee population (see definition of Ethics Training in the Glossary of Terms).
5. 2018 NPS is restated to reflect changes in survey approach. In 2019, we shortened our surveys in line with best practice and focused on more digital feedback capture instead of phone-administered surveys.
6. 2018 data was previously reported as at July 31 and has been restated as at October 31.
7. Minimum of 40% women and men in each line of business.
8. Greenhouse gas emissions data is verified by an independent third party. The verification statement is available online at: Environmental Performance.
9. Available online in June 2020 at: Environmental Performance.
10. Versus fiscal 2016, adjusting for weather and emissions factors.
11. Target to offset 90% of Scope 2 location-based emissions through the purchase of renewable energy certificates each year.
12. Amounts for 2017 and 2018 have been restated due to timing of loan closings, timing of reports and receipt of new information.
13. As at October 1.
14. As at September 30.
15. 2018 value is restated to adjust for entry-level roles (see definition of Open positions filled by internal candidates in the Glossary of Terms).

▲ 2019 data for this metric has been externally assured.

Environmental Performance: https://corporate-responsibility.bmo.com/our-practices/environmental-stewardship/enviro-targets-performances/

	Sustainability Scorecard	Access to Banking	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3
0	NaN	Number of women-owned businesses – Canada (tot...	109,3772	–	–	NaN
1	Indicator definitions are listed in the Glossa...	Loans to women-owned businesses – Canada ($ mi...	8,5103	–	–	$3 billion by June 2021
2	Terms on pages 108–111.	Business Conduct	NaN	NaN	NaN	NaN
3	NaN	Ethics training (% completion)	99.7	99.7	99.7	97%4
4	NaN	Significant fines or monetary sanctions	0	0	0	NaN
5	NaN	Contributing to Communities	NaN	NaN	NaN	NaN
6	NaN	Donations – Canada and United States ($ millions)	67.2	63.5	62.3	NaN
7	NaN	Employee volunteer hours	18,394	19,500	17,000	NaN
8	NaN	Total raised through BMO Employee Giving Campa...	22.8	22.3	21.6	NaN
9	NaN	Corporate Governance	NaN	NaN	NaN	NaN
10	NaN	Average tenure of members of Board of Director...	8	8	8	NaN
11	NaN	Women on Board of Directors (% of independent ...	35.7	35.7	41.7	No less than 1⁄3
12	NaN	Customer Experience	NaN	NaN	NaN	NaN
13	Notes	NaN	NaN	NaN	NaN	NaN
14	1. Fiscal year (November 1 to October 31), unl...	Average time to complete Banking reviews by BM...	37	37	63	NaN
15	2. As at September 30.	NaN	NaN	NaN	NaN	NaN
16	3. As at July 31.	Average time to complete Investment reviews by...	41	43	95	NaN
17	4. Based on normal employee turnover, a comple...	NaN	NaN	NaN	NaN	NaN
18	considered to include 100% of the employee pop...	Customer loyalty surveys completed	1,168,955	350,937	301,094	NaN
19	Ethics Training in the Glossary of Terms).	NaN	NaN	NaN	NaN	NaN
20	5. 2018 NPS is restated to reflect changes in ...	Banking reviews completed by BMO′ s Office of t...	350	250	247	NaN
21	shortened our surveys in line with best practi...	NaN	NaN	NaN	NaN	NaN
22	feedback capture instead of phone-administered...	Investment reviews completed by BMO′ s Office o...	52	41	56	NaN
23	6. 2018 data was previously reported as at Jul...	NaN	NaN	NaN	NaN	NaN
24	October 31.	Net Promoter Score® (NPS) – enterprise-wide	52	475	36	NaN
25	7. Minimum of 40% women and men in each line o...	NaN	NaN	NaN	NaN	NaN
26	8. Greenhouse gas emissions data is verified b...	Cyber Security, Information Security and Privacy	NaN	NaN	NaN	NaN
27	The verification statement is available online...	NaN	NaN	NaN	NaN	NaN
28	9. Available online in June 2020 at: Environme...	Substantiated complaints regarding breaches of...	4	96	6	NaN
29	10. Versus fiscal 2016, adjusting for weather ...	(total #)	NaN	NaN	NaN	NaN

After reviewing the output, we found this table was extracted correctly.

we also tried "blackrock-2020-annual-report.pdf", a report that most of the tables in it are borderless.

FINANCIAL HIGHLIGHTS

(in millions, except per share data)	2020	2019	2018	2017	2016
GAAP:					
Total revenue	**$ 16,205**	$ 14,539	$ 14,198	$ 13,600	$ 12,261
Operating income	**$ 5,695**	$ 5,551	$ 5,457	$ 5,254	$ 4,565
Operating margin	**35.1%**	38.2%	38.4%	38.6%	37.2%
Nonoperating income (expense)[1]	**$ 475**	$ 186	$ (76)	$ (32)	$ (108)
Net income attributable to BlackRock, Inc.	**$ 4,932**	$ 4,476	$ 4,305	$ 4,952	$ 3,168
Diluted earnings per common share	**$ 31.85**	$ 28.43	$ 26.58	$ 30.12	$ 19.02

(in millions, except per share data)	2020	2019	2018	2017	2016
As adjusted[2]:					
Operating income	**$ 6,284**	$ 5,551	$ 5,531	$ 5,269	$ 4,669
Operating margin	**44.9%**	43.7%	44.3%	44.1%	43.8%
Nonoperating income (expense)[1]	**$ 353**	$ 186	$ (76)	$ (32)	$ (108)
Net income attributable to BlackRock, Inc.[3]	**$ 5,237**	$ 4,484	$ 4,361	$ 3,698	$ 3,210
Diluted earnings per common share[3]	**$ 33.82**	$ 28.48	$ 26.93	$ 22.49	$ 19.27

(1) Net of net income (loss) attributable to noncontrolling interests (redeemable and nonredeemable).

(2) BlackRock reports its financial results in accordance with accounting principles generally accepted in the United States ("GAAP"); however, management believes evaluating the Company's ongoing operating results may be enhanced if investors have additional non-GAAP financial measures.

See Item 7, Management's Discussion and Analysis of Financial Condition and Results of Operations – Non-GAAP Financial Measures, for further information on non-GAAP financial measures and for as adjusted items for 2020 and 2019.

In 2018 and 2016, a restructuring charge, primarily comprised of severance and accelerated amortization expense of previously granted compensation awards, has been excluded to provide more meaningful analysis of BlackRock's ongoing operations and to ensure comparability among periods presented. In 2018, 2017 and 2016, the portion of compensation expense associated with certain long-term incentive plans funded through share distributions to participants of BlackRock stock held by PNC has been excluded because it ultimately did not impact BlackRock's book value.

(3) Net income attributable to BlackRock, Inc., as adjusted, and diluted earnings per common share, as adjusted exclude the after-tax impact of the items referred to above and exclude the effect on deferred income tax expense resulting from certain income tax matters. In 2017, $1.2 billion of net tax benefit related to The 2017 Tax Cuts and Jobs Act was excluded from net income attributable to BlackRock, Inc., as adjusted, and diluted earnings per common share, as adjusted.

ASSETS UNDER MANAGEMENT

The Company's AUM by product type for the years 2016 through 2020 is presented below.

	December 31,					
(in millions)	2020	2019	2018	2017	2016	5-Year CAGR[1]
Equity	**$ 4,419,806**	$ 3,820,329	$ 3,035,825	$ 3,371,641	$ 2,657,176	13%
Fixed income	**2,674,488**	2,315,392	1,884,417	1,855,465	1,572,365	13%
Multi-asset	**658,733**	568,121	461,884	480,278	395,007	12%
Alternatives	**235,042**	178,072	143,358	129,347	116,938	16%
Long-term	**7,988,069**	**6,881,914**	**5,525,484**	**5,836,731**	**4,741,486**	**13%**
Cash management	**666,252**	545,949	448,565	449,949	403,584	17%
Advisory	**22,359**	1,770	1,769	1,515	2,782	17%
Total	**$ 8,676,680**	**$ 7,429,633**	**$ 5,975,818**	**$ 6,288,195**	**$ 5,147,852**	**13%**

(1) Percentage represents CAGR over a five-year period (December 31, 2015 – December 31, 2020).

Component changes in AUM by product type for the five years ended December 31, 2020 are presented below.

(in millions)	December 31, 2015	Net inflows (outflows)	Acquisitions and dispositions[1]	Market change	FX impact	December 31, 2020	5-Year CAGR[2]
Equity	$ 2,423,772	$ 274,083	$ 2,590	$ 1,677,581	$ 41,780	$ 4,419,806	13%
Fixed income	1,422,368	799,393	18,539	399,477	34,711	2,674,488	13%
Multi-asset	376,336	73,572	683	196,244	11,898	658,733	12%
Alternatives	112,839	80,090	8,267	31,762	2,084	235,042	16%
Long-term	**4,335,315**	**1,227,138**	**30,079**	**2,305,064**	**90,473**	**7,988,069**	**13%**
Cash management	299,884	273,890	81,321	6,253	4,904	666,252	17%
Advisory	10,213	11,622	—	157	367	22,359	17%
Total	**$ 4,645,412**	**$ 1,512,650**	**$ 111,400**	**$ 2,311,474**	**$ 95,744**	**$ 8,676,680**	**13%**

(1) Amounts include AUM acquired in the BofA Global Capital Management transaction in April 2016, AUM acquired in the acquisition of the equity infrastructure franchise of First Reserve ("First Reserve Transaction") in June 2017, net AUM from the acquisitions of Tennenbaum Capital Partners in August 2018 ("TCP Transaction") and the asset management business of Citibanamex in September 2018 ("Citibanamex Transaction"), AUM reclassifications and net dispositions related to the transfer of BlackRock's UK Defined Contribution Administration and Platform business to Aegon N.V. in July 2018 ("Aegon Transaction"), and net AUM dispositions related to the sale of BlackRock's minority interest in DSP BlackRock Investment Managers Pvt. Ltd. to the DSP Group in August 2018 ("DSP Transaction").

(2) Percentage represents CAGR over a five-year period (December 31, 2015 – December 31, 2020).

AUM represents the broad range of financial assets managed for clients on a discretionary basis pursuant to investment management and trust agreements that are expected to continue for at least 12 months. In general, reported AUM reflects the valuation methodology that corresponds to the basis used for determining revenue (for example, net asset value). Reported AUM does not include assets for which BlackRock provides risk management or other forms of nondiscretionary advice, or assets that the Company is retained to manage on a short-term, temporary basis.

Investment management fees are typically earned as a percentage of AUM. BlackRock also earns performance fees on certain portfolios relative to an agreed-upon benchmark or return hurdle. On some products, the Company also may earn securities lending revenue. In addition, BlackRock offers its proprietary *Aladdin* investment system as well as risk management, outsourcing, advisory and other technology services, to institutional investors and wealth management intermediaries. Revenue for these services may be based on several criteria including value of positions, number of users, implementation go-lives, software solution delivery and support, and hosting services.

At December 31, 2020, total AUM was $8.68 trillion, representing a CAGR of 13% over the last five years. AUM growth during the period was achieved through the combination of net market valuation gains, net inflows and acquisitions, including BofA Global Capital Management, which added $80.6 billion of AUM in 2016, the First Reserve Transaction, which added $3.3 billion of AUM in 2017 and the net AUM impact from the TCP Transaction, the Citibanamex Transaction, the Aegon Transaction and the DSP Transaction, which added $27.5 billion of AUM in 2018. Our AUM mix encompasses a broadly diversified product range, as described below.

The Company considers the categorization of its AUM by client type, product type, investment style, and client region useful to understanding its business. The following discussion of the Company's AUM will be organized as follows:

Client Type	Product Type	Investment Style	Client Region
• Retail	• Equity	• Active	• Americas
• *i*Shares ETFs	• Fixed Income	• Index and *i*Shares ETFs	• Europe, the Middle East and Africa ("EMEA")
• Institutional	• Multi-asset		• Asia-Pacific
	• Alternatives		
	• Cash Management		

CLIENT TYPE

BlackRock serves a diverse mix of institutional and retail clients across the globe, with a regionally focused business model. BlackRock leverages the benefits of scale across global investment, risk and technology platforms while at the same time using local distribution presence to deliver solutions for clients. Furthermore, our structure facilitates strong teamwork globally across both functions and regions in order to enhance our ability to leverage best practices to serve our clients and continue to develop our talent.

Clients include tax-exempt institutions, such as defined benefit and defined contribution pension plans, charities, foundations and endowments; official institutions, such as central banks, sovereign wealth funds, supranationals and other government entities; taxable institutions, including insurance companies, financial institutions, corporations and third-party fund sponsors, and retail intermediaries.

*i*Shares ETFs are a growing component of both institutional and retail client portfolios. However, as *i*Shares ETFs are traded on exchanges, complete transparency on the ultimate end-client is unavailable. Therefore, *i*Shares ETFs are presented as a separate client type below, with investments in *i*Shares ETFs by institutions and retail clients excluded from figures and discussions in their respective sections.

Financial Highlights

Please review the Important Notes on page 25 for information on certain non-GAAP figures shown through page 24, as well as for source information on other data points on pages 2 through 24.

(in millions)	2020	2019	2018
Total AUM (end of period)	$ 8,676,680	$ 7,429,633	$ 5,975,818
Revenue	16,205	14,539	14,198
Operating income, as adjusted	6,284	5,551	5,531
Operating margin, as adjusted	44.9%	43.7%	44.3%
Net income attributable to BLK, GAAP	4,932	4,476	4,305
Net income attributable to BLK, as adjusted	5,237	4,484	4,361
Diluted weighted-average common shares	154.8	157.5	161.9
Per share			
Diluted earnings, GAAP	$ 31.85	$ 28.43	$ 26.58
Diluted earnings, as adjusted	33.82	28.48	26.93
Dividends declared	14.52	13.20	12.02

Important Notes

Opinions

Opinions expressed through page 24 are those of BlackRock, Inc. as of March 2021 and are subject to change. Investment involves risk including the loss of principal. The companies mentioned in this document are not meant to be a recommendation to buy or sell any security.

BlackRock data points

All data through page 24 reflects as-adjusted full-year 2020 results or is as of December 31, 2020, unless otherwise noted. 2020 organic growth is defined as full-year 2020 net flows divided by assets under management (AUM) for the entire firm, a particular segment or particular product as of December 31, 2019. Long-term product offerings include active and passive strategies across equity, fixed income, multi-asset and alternatives, and exclude AUM and flows from the cash management and advisory businesses.

Industry data points

All data is as of December 31, 2020 unless otherwise noted.

As of December 31, 2020, the SNL U.S. Asset Manager Index included: Affiliated Managers Group Inc.; AllianceBernstein Holding L.P.; Ameriprise Financial Inc.; Apollo Global Management Inc.; Ares Management Corporation; Artisan Partners Asset Management Inc.; Ashford Inc.; Associated Capital Group Inc.; BlackRock, Inc.; Blackstone Group Inc.; BrightSphere Investment Group Inc.; Carlyle Group Inc.; Cohen & Steers Inc.; Diamond Hill Investment Group; Federated Hermes Inc.; Fifth Street Asset Management Inc.; Franklin Resources Inc.; GAMCO Investors Inc.; Great Elm Group; Hamilton Lane Inc.; Hennessy Advisors Inc.; Invesco Ltd.; Janus Henderson Group Plc.; KKR & Co; Manning & Napier Inc.; Medley Management Inc.; Pzena Investment Management Inc.; Safeguard Scientifics Inc.; Sculptor Capital Management Inc.; SEI Investments Co.; Silvercrest Asset Management Group; T. Rowe Price Group Inc.; The Gabelli Equity Trust; U.S. Global Investors Inc.; Victory Capital Holdings Inc.; Virtus Investment Partners Inc.; Waddell & Reed Financial Inc.; Westwood Holdings Group Inc.; WisdomTree Investments Inc.

GAAP and as-adjusted results

See pages 40–41 of our 2020 10-K for an explanation of the use of non-GAAP financial measures and a reconciliation to GAAP.

Performance notes

Past performance is not indicative of future results. Except as specified, the performance information shown is as of December 31, 2020 and is based on preliminary data available at that time. The performance data shown reflects information for all actively and passively managed equity and fixed income accounts, including U.S. registered investment companies, European-domiciled retail funds and separate accounts for which performance data is available, including performance data for high net worth accounts available as of November 30, 2020. The performance data does not include accounts terminated prior to December 31, 2020 and accounts for which data has not yet been verified. If such accounts had been included, the performance data provided may have substantially differed from that shown.

Performance comparisons shown are gross-of-fees for institutional and high net worth separate accounts, and net-of-fees for retail funds. The performance tracking shown for index accounts is based on gross-of-fees performance and includes all institutional accounts and all *iShares* funds globally using an index strategy. AUM information is based on AUM available as of December 31, 2020 for each account or fund in the asset class shown without adjustment for overlapping management of the same account or fund. Fund performance reflects the reinvestment of dividends and distributions.

Performance shown is derived from applicable benchmarks or peer median information, as selected by BlackRock, Inc. Peer medians are based in part on data either from Lipper, Inc. or Morningstar, Inc. for each included product.

	Net income attributable to BLK, GAAP	Unnamed: 0	4,932	Unnamed: 1	4,476	Unnamed: 2	4,305	organic growth is defined as full-	Inc.; Safeguard Scientifics Inc.; Sculptor Capital Management Inc.;	Unnamed: 3	differed from that shown.
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	year 2020 net flows divided by	SEI Investments Co.; Silvercrest	NaN	Performance comparisons shown are
1	Net income attributable to BLK, as adjusted	NaN	5,237	NaN	4,484	NaN	4,361	assets under management (AUM)	Asset Management Group; T. Rowe	NaN	gross-of-fees for institutional and
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	for the entire firm, a particular	Price Group Inc.; The Gabelli	NaN	high net worth separate accounts,
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	segment or particular product	Equity Trust; U.S. Global Investors	NaN	and net-of-fees for retail funds. The
4	Diluted weighted-average common shares	NaN	154.8	NaN	157.5	NaN	161.9	as of December 31, 2019. Long-	Inc.; Victory Capital Holdings Inc.;	NaN	performance tracking shown for
5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	term product offerings include	Virtus Investment Partners Inc.;	NaN	index accounts is based on gross-
6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	active and passive strategies	Waddell & Reed Financial Inc.;	NaN	of-fees performance and includes
7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	across equity, fixed income,	Westwood Holdings Group Inc.;	NaN	all institutional accounts and all
8	Per share	NaN	NaN	NaN	NaN	NaN	NaN	multi-asset and alternatives, and	WisdomTree Investments Inc.	NaN	iShares funds globally using an
9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	exclude AUM and flows from the	NaN	NaN	index strategy. AUM information
10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	cash management and advisory	NaN	NaN	is based on AUM available as of
11	Diluted earnings, GAAP	$	31.85	$	28.43	$	26.58	businesses.	GAAP and as-	NaN	December 31, 2020 for each account

In that case, Tabula can’t detect tables correctly, it always mistakes too much other text for tables. Even worse, it can’t extract content from those clear borderless table.

@aegis4048 has mentioned that there exist two modes of table extraction in tabula. https://aegis4048.github.io/parse-pdf-files-while-retaining-structure-with-tabula-py

lattice=True forces PDFs to be extracted using lattice-mode extraction. It recognizes each cells based on ruling lines, or borders of each cell.

and

stream=True forces PDFs to be extracted using stream-mode extraction. This mode is used when there are no ruling lines to differentiate one cell from the other. Instead, it uses spacings among each cells to recognize each cell.

As it describes, lattice mode recognizes tables based on ruling lines or borders, as a result, it doesn't perform well on borderless table. In the article he has mentioned that we can use stream mode to extract borderless table. It may work when extracting a pdf with only one table, but this does not apply to our case. Since the tables are embedded in the paragraph, there are spacings everywhere, between paragraph and paragraph, paragraph and table, or table and table. In that case, tabula won't be able to detect tables from the paragraphs, and it will simply extract all the text as tables.

There are also some other text-based table extraction tools such as Camelot, but they have the same problem as tabula: the tables cannot be accurately detected from the paragraphs.

Since one single table extraction tool cannot detect tables well,
can we detect them first and then extract them separately?

BMO's Purpose Manifesto

Since the beginning, BMO has had a purpose beyond profit to be a champion for progress and a catalyst for change. We are a partner for our customers seeking to fulfill their dreams, we are an employer for the tens of thousands of people who work with us every day, and we are proud members of the communities in which we live and work. Our commitment to driving positive change for our customers, employees and communities has spanned more than 200 years. In short, our Purpose is to Boldly Grow the Good *in business and life*.

Our success has involved adapting to a changing society while helping our customers do the same. Today, from human rights to climate change to sustainable finance, the challenges we face are increasingly complex and interconnected. In response, we are strengthening our commitment to sustainability and to navigating these changes together with our stakeholders.

We believe that profit and progress go hand in hand. We will think big and pursue our growth objectives confidently, balancing risk and opportunity to create value that is shared with the world around us. Our Purpose encourages us to do the right thing every day.

In support of our Purpose, BMO has committed to **double the good** by 2025

- **For a sustainable future:** we will mobilize $400 billion for sustainable finance, and launch an impact investment fund to find and scale solutions to sustainability challenges.
- **For a thriving economy:** we will double BMO's support for small businesses and for women entrepreneurs.
- **For an inclusive society:** we will reduce barriers to inclusion for employees, customers and stakeholders partnering with BMO.

Inspired by the United Nations Sustainable Development Goals (SDGs) and grounded in BMO's values of integrity, empathy, diversity and responsibility, these commitments will shape everything we do: the way we serve our customers and the products and services we design for them; our ongoing hiring practices and employee training and development programs; and our community giving and engagement with our partners.

To deliver on and amplify our Purpose, we will:

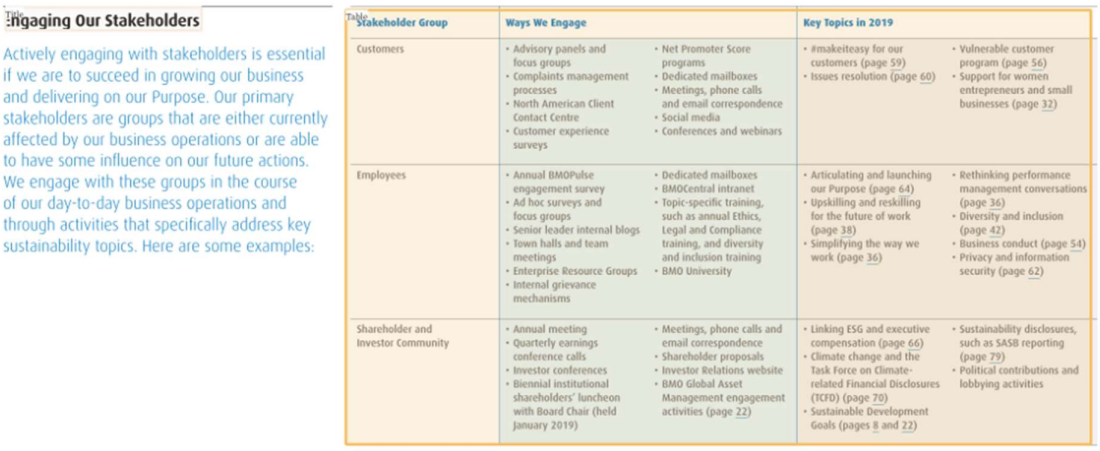
- Work to increase positive impacts and reduce negative impacts across BMO's operations and value chain.
- Develop new products and services to help our customers achieve society's shared economic, social and environmental objectives.
- Embed sustainability more deeply within our operations and internal procedures.
- Support collaboration, learning and innovation through our community giving activities.
- Actively engage stakeholders and investee companies to drive best practices.

BOLDLY GROW THE GOOD
IN BUSINESS AND LIFE

We will continue to focus our efforts in areas where we can have a measurable impact on

- **Our shareholders:** generating long-term value for our shareholders and engaging with them openly.
- **Our customers:** fostering inclusive relationships with our customers, including women entrepreneurs and small businesses, to address the challenge of economic disparity.
- **Our employees:** championing diversity and inclusion across our organization and preparing our colleagues for the jobs of the future.
- **Our environment:** managing our environmental footprint and financing and advising our clients during the transition to a resource-efficient, low-carbon and circular economy.
- **Our community:** building stronger, healthier communities through donations, sponsorships and volunteering.
- **Our suppliers:** dealing fairly and ethically with our suppliers and working with them to improve their practices.
- **Our peers:** collaborating on thought leadership and innovation initiatives and leveraging our combined influence to create positive and sustainable change.

Together, BMO's operating groups are mobilizing to make real progress toward these goals. By working together, we can all elevate our performance and be the change that grows the good for us all.

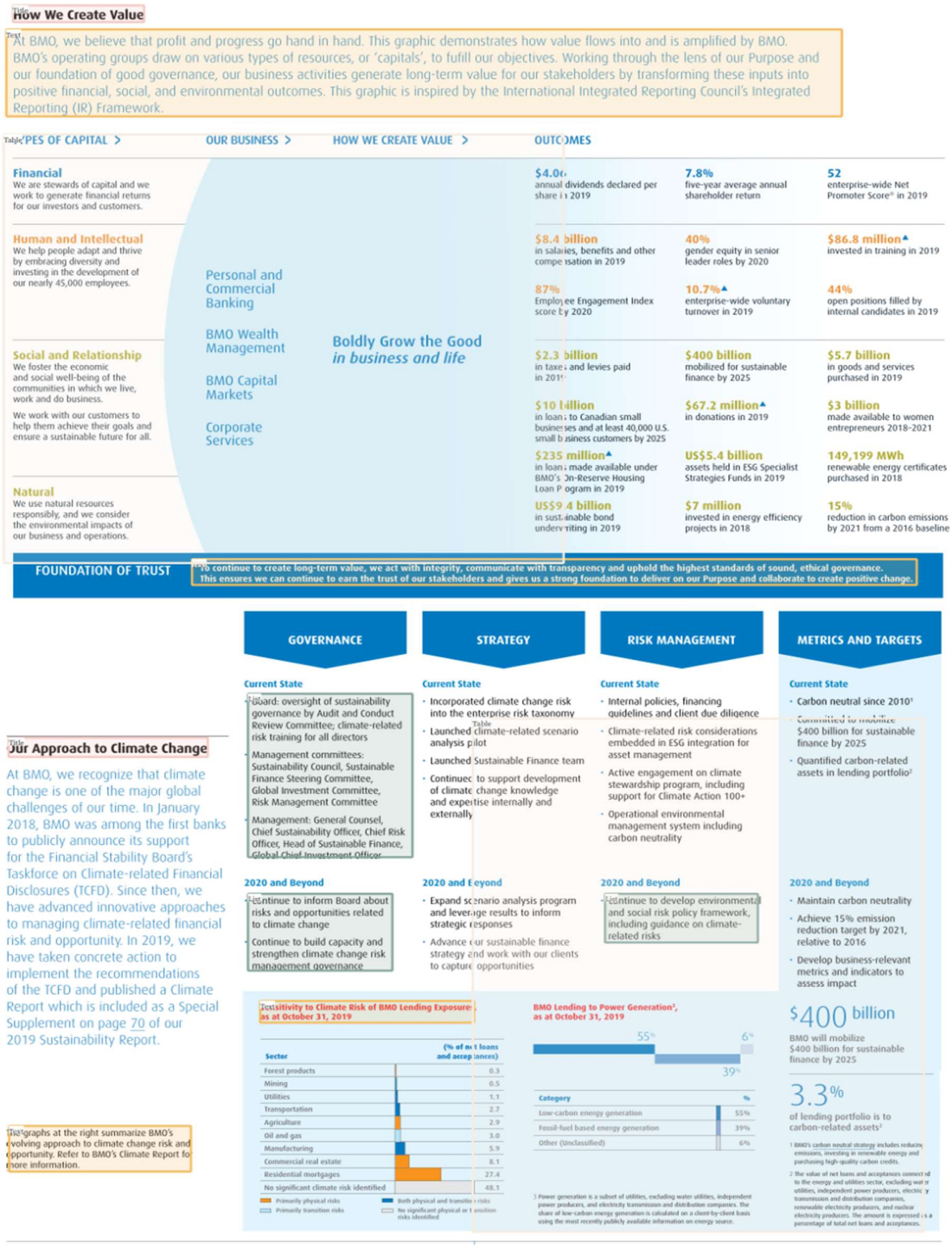


Engaging Our Stakeholders

Actively engaging with stakeholders is essential if we are to succeed in growing our business and delivering on our Purpose. Our primary stakeholders are groups that are either currently affected by our business operations or are able to have some influence on our future actions. We engage with these groups in the course of our day-to-day business operations and through activities that specifically address key sustainability topics. Here are some examples:

Stakeholder Group	Ways We Engage		Key Topics in 2019	
Customers	• Advisory panels and focus groups • Complaints management processes • North American Client Contact Centre • Customer experience surveys	• Net Promoter Score programs • Dedicated mailboxes • Meetings, phone calls and email correspondence • Social media • Conferences and webinars	• #makeiteasy for our customers (page 59) • Issues resolution (page 60)	• Vulnerable customer program (page 56) • Support for women entrepreneurs and small businesses (page 32)
Employees	• Annual BMOPulse engagement survey • Ad hoc surveys and focus groups • Senior leader internal blogs • Town halls and team meetings • Enterprise Resource Groups • Internal grievance mechanisms	• Dedicated mailboxes • BMOCentral intranet • Topic-specific training, such as annual Ethics, Legal and Compliance training, and diversity and inclusion training • BMO University	• Articulating and launching our Purpose (page 64) • Upskilling and reskilling for the future of work (page 38) • Simplifying the way we work (page 36)	• Rethinking performance management conversations (page 36) • Diversity and inclusion (page 42) • Business conduct (page 54) • Privacy and information security (page 62)
Shareholder and Investor Community	• Annual meeting • Quarterly earnings conference calls • Investor conferences • Biennial institutional shareholders' luncheon with Board Chair (held January 2019)	• Meetings, phone calls and email correspondence • Shareholder proposals • Investor Relations website • BMO Global Asset Management engagement activities (page 22)	• Linking ESG and executive compensation (page 66) • Climate change and the Task Force on Climate-related Financial Disclosures (TCFD) (page 70) • Sustainable Development Goals (pages 8 and 22)	• Sustainability disclosures, such as SASB reporting (page 79) • Political contributions and lobbying activities

We use OCR to detect the tables in the first 20 pages, and we can see that where tabula make mistakes, OCR can correctly identify them; where tables are clear, OCR won't make errors.



How We Create Value

At BMO, we believe that profit and progress go hand in hand. This graphic demonstrates how value flows into and is amplified by BMO. BMO's operating groups draw on various types of resources, or 'capitals', to fulfill our objectives. Working through the lens of our Purpose and our foundation of good governance, our business activities generate long-term value for our stakeholders by transforming these inputs into positive financial, social, and environmental outcomes. This graphic is inspired by the International Integrated Reporting Council's Integrated Reporting (IR) Framework.

TYPES OF CAPITAL > | OUR BUSINESS > | HOW WE CREATE VALUE > | OUTCOMES

Financial
We are stewards of capital and we work to generate financial returns for our investors and customers.

Human and Intellectual
We help people adapt and thrive by embracing diversity and investing in the development of our nearly 45,000 employees.

Social and Relationship
We foster the economic and social well-being of the communities in which we live, work and do business.

We work with our customers to help them achieve their goals and ensure a sustainable future for all.

Natural
We use natural resources responsibly, and we consider the environmental impacts of our business and operations.

Personal and Commercial Banking
BMO Wealth Management
BMO Capital Markets
Corporate Services

Boldly Grow the Good in business and life

$4.0[illegible] annual dividends declared per share in 2019
7.8% five-year average annual shareholder return
52 enterprise-wide Net Promoter Score in 2019
$8.4 billion in salaries, benefits and other compensation in 2019
40% gender equity in senior leader roles by 2020
$86.8 million▲ invested in training in 2019
87% Employee Engagement Index score by 2020
10.7%▲ enterprise-wide voluntary turnover in 2019
44% open positions filled by internal candidates in 2019
$2.3 billion in taxes and levies paid in 2019
$400 billion mobilized for sustainable finance by 2025
$5.7 billion in goods and services purchased in 2019
$10 billion in loans to Canadian small businesses and at least 40,000 U.S. small business customers by 2025
$67.2 million▲ in donations in 2019
$3 billion made available to women entrepreneurs 2018-2021
$235 million▲ in loans made available under BMO's On-Reserve Housing Loan Program in 2019
US$5.4 billion assets held in ESG Specialist Strategies Funds in 2019
149,199 MWh renewable energy certificates purchased in 2018
US$9.4 billion in sustainable bond underwriting in 2019
$7 million invested in energy efficiency projects in 2018
15% reduction in carbon emissions by 2021 from a 2016 baseline

FOUNDATION OF TRUST
To continue to create long-term value, we act with integrity, communicate with transparency and uphold the highest standards of sound, ethical governance. This ensures we can continue to earn the trust of our stakeholders and gives us a strong foundation to deliver on our Purpose and collaborate to create positive change.

Our Approach to Climate Change

At BMO, we recognize that climate change is one of the major global challenges of our time. In January 2018, BMO was among the first banks to publicly announce its support for the Financial Stability Board's Taskforce on Climate-related Financial Disclosures (TCFD). Since then, we have advanced innovative approaches to managing climate-related financial risk and opportunity. In 2019, we have taken concrete action to implement the recommendations of the TCFD and published a Climate Report which is included as a Special Supplement on page 70 of our 2019 Sustainability Report.

The graphs at the right summarize BMO's evolving approach to climate change risk and opportunity. Refer to BMO's Climate Report for more information.

GOVERNANCE

Current State
- Board: oversight of sustainability governance by Audit and Conduct Review Committee; climate-related risk training for all directors
- Management committees: Sustainability Council, Sustainable Finance Steering Committee, Global Investment Committee, Risk Management Committee
- Management: General Counsel, Chief Sustainability Officer, Chief Risk Officer, Head of Sustainable Finance, Global Chief Investment Officer

2020 and Beyond
- Continue to inform Board about risks and opportunities related to climate change
- Continue to build capacity and strengthen climate change risk management governance

STRATEGY

Current State
- Incorporated climate change risk into the enterprise risk taxonomy
- Launched climate-related scenario analysis pilot
- Launched Sustainable Finance team
- Continued to support development of climate change knowledge and expertise internally and externally

2020 and Beyond
- Expand scenario analysis program and leverage results to inform strategic responses
- Advance our sustainable finance strategy and work with our clients to capture opportunities

RISK MANAGEMENT

Current State
- Internal policies, financing guidelines and client due diligence
- Climate-related risk considerations embedded in ESG integration for asset management
- Active engagement on climate stewardship program, including support for Climate Action 100+
- Operational environmental management system including carbon neutrality

2020 and Beyond
- Continue to develop environmental and social risk policy framework, including guidance on climate-related risks

METRICS AND TARGETS

Current State
- Carbon neutral since 2010[1]
- Committed to mobilize $400 billion for sustainable finance by 2025
- Quantified carbon-related assets in lending portfolio[2]

2020 and Beyond
- Maintain carbon neutrality
- Achieve 15% emission reduction target by 2021, relative to 2016
- Develop business-relevant metrics and indicators to assess impact

Sensitivity to Climate Risk of BMO Lending Exposure, as at October 31, 2019

Sector	(% of net loans and acceptances)
Forest products	0.3
Mining	0.5
Utilities	5.1
Transportation	2.7
Agriculture	2.9
Oil and gas	3.0
Manufacturing	5.9
Commercial real estate	8.1
Residential mortgages	27.4
No significant climate risk identified	48.1

Primarily physical risks; Primarily transition risks; Both physical and transition risks; No significant physical or transition risks identified

BMO Lending to Power Generation[3], as at October 31, 2019

55% | 6% | 39%

Category	%
Low-carbon energy generation	55%
Fossil-fuel based energy generation	39%
Other (Unclassified)	6%

3 Power generation is a subset of utilities, excluding water utilities, independent power producers, and electricity transmission and distribution companies. The share of low-carbon energy generation is calculated on a client-by-client basis using the most recently publicly available information on energy source.

$400 billion
BMO will mobilize $400 billion for sustainable finance by 2025

3.3%
of lending portfolio is to carbon-related assets[2]

1 BMO's carbon neutral strategy includes reducing emissions, investing in renewable energy and purchasing high-quality carbon credits.

2 The value of net loans and acceptances connected to the energy and utilities sector, excluding water utilities, independent power producers, electricity transmission and distribution companies, renewable electricity producers, and nuclear electricity producers. The amount is expressed as a percentage of total net loans and acceptances.

Although the accuracy rate is not 100%, OCR still significantly reduces the error rate, which will greatly reduce our efforts to eliminate useless information.

Therefore, I decided to try to crop out the tables first and then extract them using a different OCR method.

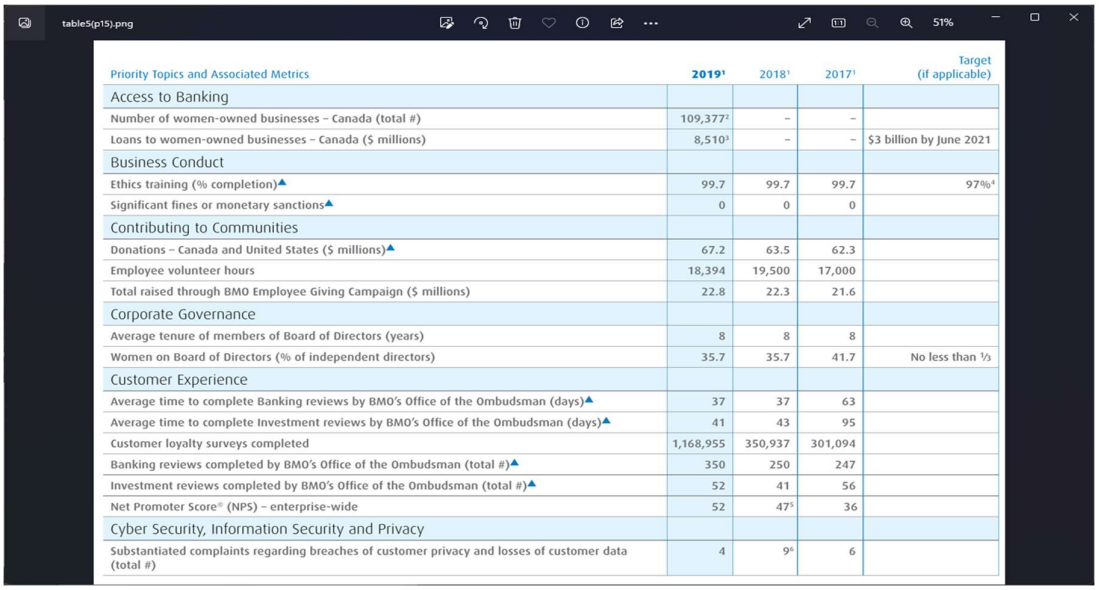
- Why OCR method?
  Tabula and Camelot only work with text-based pdfs, even if we try to convert the cropped image to an image-based pdf file, they can't extract it.

- Why a different OCR method?
  The method layoutparser[ocr] didn't work well in extracting table, so we are looking for a more powerful tool.

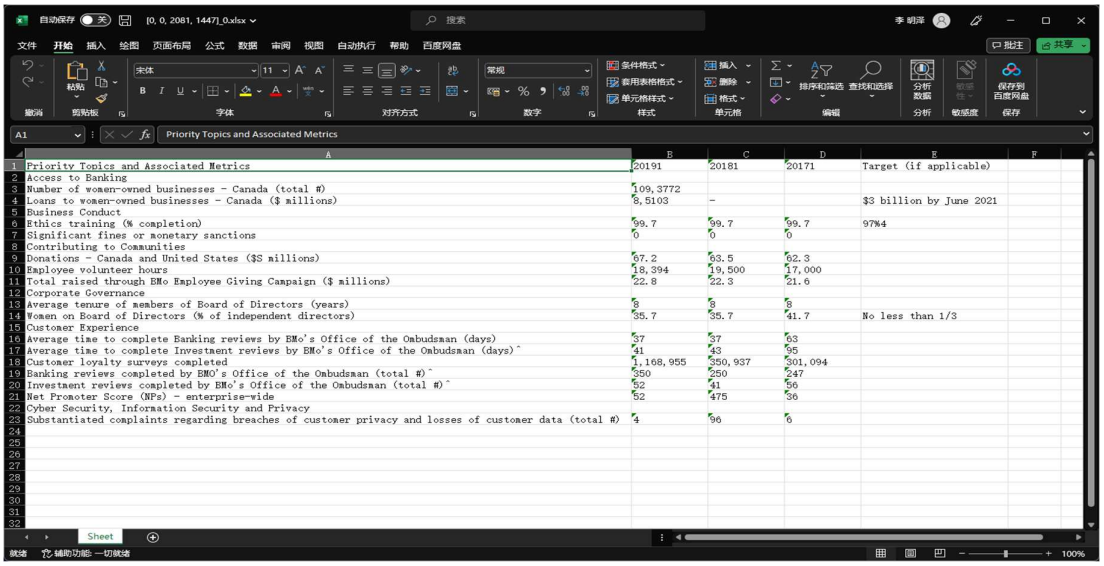
Fortunately, we found PP-Structure, an intelligent document analysis system developed by the PaddleOCR team, which aims to help developers better complete tasks related to document understanding such as layout analysis and table recognition.

More information:
https://github.com/PaddlePaddle/PaddleOCR/blob/dygraph/ppstructure/README.md

Here we mainly use its table recognition function.



Priority Topics and Associated Metrics	2019[1]	2018[1]	2017[1]	Target (if applicable)
Access to Banking				
Number of women-owned businesses – Canada (total #)	109,377[2]	–	–	
Loans to women-owned businesses – Canada ($ millions)	8,510[3]	–	–	$3 billion by June 2021
Business Conduct				
Ethics training (% completion)▲	99.7	99.7	99.7	97%[4]
Significant fines or monetary sanctions▲	0	0	0	
Contributing to Communities				
Donations – Canada and United States ($ millions)▲	67.2	63.5	62.3	
Employee volunteer hours	18,394	19,500	17,000	
Total raised through BMO Employee Giving Campaign ($ millions)	22.8	22.3	21.6	
Corporate Governance				
Average tenure of members of Board of Directors (years)	8	8	8	
Women on Board of Directors (% of independent directors)	35.7	35.7	41.7	No less than ⅓
Customer Experience				
Average time to complete Banking reviews by BMO's Office of the Ombudsman (days)▲	37	37	63	
Average time to complete Investment reviews by BMO's Office of the Ombudsman (days)▲	41	43	95	
Customer loyalty surveys completed	1,168,955	350,937	301,094	
Banking reviews completed by BMO's Office of the Ombudsman (total #)▲	350	250	247	
Investment reviews completed by BMO's Office of the Ombudsman (total #)▲	52	41	56	
Net Promoter Score® (NPS) – enterprise-wide	52	47[5]	36	
Cyber Security, Information Security and Privacy				
Substantiated complaints regarding breaches of customer privacy and losses of customer data (total #)	4	9[6]	6	



	A	B	C	D	E
1	Priority Topics and Associated Metrics	20191	20181	20171	Target (if applicable)
2	Access to Banking				
3	Number of women-owned businesses – Canada (total #)	109,3772			
4	Loans to women-owned businesses – Canada ($ millions)	8,5103	–		$3 billion by June 2021
5	Business Conduct				
6	Ethics training (% completion)	99.7	99.7	99.7	97%4
7	Significant fines or monetary sanctions	0	0	0	
8	Contributing to Communities				
9	Donations – Canada and United States ($S millions)	67.2	63.5	62.3	
10	Employee volunteer hours	18,394	19,500	17,000	
11	Total raised through BMo Employee Giving Campaign ($ millions)	22.8	22.3	21.6	
12	Corporate Governance				
13	Average tenure of members of Board of Directors (years)	8	8	8	
14	Women on Board of Directors (% of independent directors)	35.7	35.7	41.7	No less than 1/3
15	Customer Experience				
16	Average time to complete Banking reviews by BMo's Office of the Ombudsman (days)	37	37	63	
17	Average time to complete Investment reviews by BMo's Office of the Ombudsman (days)^	41	43	95	
18	Customer loyalty surveys completed	1,168,955	350,937	301,094	
19	Banking reviews completed by BMO's Office of the Ombudsman (total #)^	350	250	247	
20	Investment reviews completed by BMo's Office of the Ombudsman (total #)^	52	41	56	
21	Net Promoter Score (NPs) – enterprise-wide	52	475	36	
22	Cyber Security, Information Security and Privacy				
23	Substantiated complaints regarding breaches of customer privacy and losses of customer data (total #)	4	96	6	

As we can see, it extracted the cropped table correctly, and it recognized the header as well.

table2(p16).png

(in millions)	December 31, 2015	Net inflows (outflows)	Acquisitions and dispositions(1)	Market change	FX impact	December 31, 2020	5-Year CAGR(2)
Equity	$ 2,423,772	$ 274,083	$ 2,590	$ 1,677,581	$ 41,780	$ 4,419,806	13%
Fixed income	1,422,368	799,393	18,539	399,477	34,711	2,674,488	13%
Multi-asset	376,336	73,572	683	196,244	11,898	658,733	12%
Alternatives	112,839	80,090	8,267	31,762	2,084	235,042	16%
Long-term	**4,335,315**	**1,227,138**	**30,079**	**2,305,064**	**90,473**	**7,988,069**	**13%**
Cash management	299,884	273,890	81,321	6,253	4,904	666,252	17%
Advisory	10,213	11,622	—	157	367	22,359	17%
Total	**$ 4,645,412**	**$ 1,512,650**	**$ 111,400**	**$ 2,311,474**	**$ 95,744**	**$ 8,676,680**	**13%**

[0, 0, 2044, 582]_0.xlsx

(in millions)	December 31, 2015	Net inflows (outflows)	Acquisitions and dispositions(1)	Market change	FX impact	December 31, 2020	5-Year CAGR(2)
Equity	$ 2,423,772	$ 274,083	$ 2,590	$ 1,677,581	$ 41,780	$ 4,419,806	13%
Fixed income	1,422,368	799,393	18,539	399,477	34,711	2,674,488	13%
Multi-asset	376,336	73,572	683	196,244	11,898	658,733	12%
Alternatives	112,839	80,090	8,267	31,762	2,084	235,042	16%
Long-term	4,335,315	1,227,138	30,079	2,305,064	90,473	7,988,069	13%
Cash management	299,884	273,890	81,321	6,253	4,904	666,252	17%
Advisory	10,213	11,622		157	367	22,359	17%
Total	$ 4,645,412	$ 1,512,650	$ 111,400	$ 2,311,474	$ 95,744	$ 8,676,680	13%

For those borderless table, it can also extract them.

Therefore, our *TABLE_RXTRACT.py* will first detect the table with layoutparser, using the coordinates to crop the table, and then extract the cropped image with PaddleOCR.

Complement:

1. If layoutparser does not detect all the tables correctly, we can manually crop the tables and enter them into *image_extract.py,* it will also extract this table in excel form.
2. It is possible to detect plots and graphs in the same way. Layoutparser also support figure detection, we can get their coordinates and crop them out. However, the model we currently apply does not have a high accuracy rate in figure detection. A more specialized model is needed if we want to implement figure extraction.